Attorney Docket No. NETAP021
Client Docket No. 01.166.01

APPLICATION FOR UNITED STATES PATENT

SPAM REPORT GENERATION SYSTEM AND METHOD

By Inventors:

Luke David Jagger
3 Egypt Way
Aylesbury
Bucks
HP19 8GP
United Kingdom
Citizen of United Kingdom

William R. Dennis
3 Nuthatch
Aylesbury
Buckinghamshire
United Kingdom
HP190WF
Citizen of United Kingdom

Anton Christian Rothwell
7 Parmiter Close
Aylesbury
Bucks
HP19 8GS
United Kingdom
Citizen of United Kingdom

10072708.020502

Assignee: **Networks Associates Technology, Inc.**
3965 Freedom Circle
Santa Clara, CA 95054

Entity: Large

RITTER, LANG & KAPLAN LLP
12930 Saratoga Ave., Suite D1
Saratoga, CA 95070
(408) 446-8690

SPAM REPORT GENERATION SYSTEM AND METHOD

BACKGROUND OF THE INVENTION

The present invention relates generally to electronic mail ('e-mail'), and more specifically, to a method and system for identifying and reporting SPAM e-mail messages.

Unsolicited bulk e-mail, commonly referred to as "SPAM", is increasingly becoming a nuisance to computer users. SPAM itself is not illegal, however, the content of some messages may violate laws or the SPAM initiator's contract with his Internet Service Provider (ISP). SPAM e-mail is generally defined as an unsolicited mailing, usually to a large number of people. SPAM can be very annoying to the recipient because it interrupts other activities, consumes system resources, and requires active efforts by recipients who want to dispose of these unwanted messages.

SPAM is also an increasing problem for Internet service providers and entities with easily identifiable e-mail addresses such as large corporations. ISPs object to junk mail because it reduces their users' satisfaction of the services. Corporations want to eliminate junk mail because it reduces worker productivity.

SPAM impacts organizations by occupying employees' time and increasing security risks. Time is spent by employees to open each message, classify it as legitimate or junk e-mail, and delete the message. Time may also be spent by employees following up on advertising content while on the job. Employees may also be deceived into acting improperly, such as to release confidential information, due to a forged message. There is also a loss of the network administrator's time in dealing with SPAM and forged messages, as well as the use of network bandwidth, disk space, and system memory required to store the messages. Also, in the process of deleting junk mail, users may inadvertently discard or overlook other important messages. Another objection to SPAM is that it is frequently used to advertise objectionable, fraudulent, or dangerous content, such as pornography or to propagate financial scams such as illegal pyramid schemes.

The person or organization that generates the junk mail (referred to as a 'spammer') often gets around filtering methods by using a different e-mail address for each mailing or forwarding his e-mail by way of an intermediary to conceal the actual origin. Instead of mailing directly from an easily traced account at a major Internet service provider, spammers may, for instance, send their e-mail from a SPAM-friendly network, using forged headers, and relay the message through intermediate hosts. However, the e-mail message often contains an actual web site that relates to the message so that the recipient can find additional information on the advertised product or service. No action can be taken against the person or organization that generates the junk mail unless that person or organization is identified and someone reports the problem to the relevant authority.

There is, therefore, a need for a system and method for identifying and reporting SPAM to the appropriate authority so that the authority can take action to prevent the spammer from distributing further unsolicited e-mail.

10072708.020502

SUMMARY OF THE INVENTION

A method and system for generating a report on an unsolicited electronic message and sending the report to the relevant authority are disclosed.

A method of the present invention generally comprises receiving an electronic mail message and determining whether the electronic message is an unsolicited message. If the message is an unsolicited message, it is examined to identify a network address relating to the message and an authority hosting the network address. A report is then generated containing the identified network address and the hosting authority.

The generated report is sent to the hosting authority or to a central managed service provider that collects reports and transmits them to the appropriate authority. The reports may also be held and collected over a period of time before they are sent out.

A system of the present invention generally comprises a detector operable to detect a network address within an electronic message identified as an unsolicited message and a host identifier operable to identify an authority hosting the network. The system further includes a report generator operable to generate a report containing the identified network address and hosting authority and a storage medium configured to at least temporarily store the identified network address and hosting authority.

In one embodiment, the system includes a database that contains common words and phrases that can be used in searching for a URL within the message. The host identifier may then use an Internet tool to identify the organization hosting the web site of the URL.

10072708.020502

In another aspect of the present invention, a computer product generally comprises code that receives an electronic mail message and determines whether the electronic message is an unsolicited message. The product further includes code that examines the message to identify a network address relating to the message if the message is an unsolicited message and code that identifies an authority hosting the network address and generates a report containing the identified network address. A computer readable medium is provided to store the computer codes.

The above is a brief description of some deficiencies in the prior art and advantages of the present invention. Other features, advantages, and embodiments of the invention will be apparent to those skilled in the art from the following description, drawings, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram illustrating a network utilizing a system and method of the present invention.

Fig. 2 is a block diagram illustrating a SPAM reporting system of the present invention.

Fig. 3 is a block diagram of a computer system that can be utilized to execute software of an embodiment of the invention.

Fig. 4 is a flowchart illustrating a process of the present invention for generating a SPAM report

Corresponding reference characters indicate corresponding parts throughout the several views of the drawings.

10072708.020502

DETAILED DESCRIPTION OF THE INVENTION

The following description is presented to enable one of ordinary skill in the art to make and use the invention. Descriptions of specific embodiments and applications are provided only as examples and various modifications will be readily apparent to those skilled in the art. The general principles described herein may be applied to other embodiments and applications without departing from the scope of the invention. Thus, the present invention is not to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features described herein. For purpose of clarity, details relating to technical material that is known in the technical fields related to the invention have not been described in detail.

The present invention provides a method and system for generating a report upon detection of unsolicited or SPAM electronic mail ('e-mail') messages. The report is preferably automatically generated upon detection of an unsolicited e-mail. In one embodiment, the system sends the report to the relevant authority (e.g., Internet Service Provider (ISP) or backbone provider hosting the spammer). The report may be used by the authority to take action if necessary in shutting down a spammer's web site and mail access.

Referring now to the drawings, and first to Fig. 1, a system that may utilize the present invention is shown and generally indicated at 20. The system 20 includes user computers 22, 24, 26, 28 (described further below) in communication with one another through a network. User computers 22, 24, 26 may receive, for example, unsolicited e-mail from user computers 28 which are operated by spammers. User computers 22 receive e-mail through service provider 30 and user computer 26 receives e-mail through gateway 32. Computer 24 receives e-mail directly from the network. SPAM reporting systems (SRS) 40 are installed within the network and configured to detect SPAM, identify a source host, and generate a report transmitting information about the SPAM to the source host. Spam reporting system 40 is preferably installed at an Internet service provider or gateway, but may also be installed at a user computer.

In the network shown in Fig. 1, service provider 30, gateway 32, and computer 24 each include SPAM reporting system 40. In the present example, user computers 28 are coupled to a service provider 34 along with server 36 which includes a web site set up by the spammers of user computers 28. It is to be understood that the SPAM may be sent from a different service provider than the one which hosts the spammer's web site. Upon receiving SPAM at computers 22, 24, or 26 from computers 28, a network address of the web site located at server 36 is identified, the service provider 34 supporting the web site is identified, and a report is generated by the SPAM reporting system 40, as

10072703.020502

described in detail below. The report contains details on the SPAM generated by service provider's customer and is transmitted to the service provider 34 for investigation by the appropriate authorities.

The network may include any number of servers 36 for hosting network sites (web sites). The servers are typically connected to the network at points of presence (POPs), established by network service providers at a variety of geographic locations. A given geographic location, such as a metropolitan area, will typically contain multiple POPs established by different network service providers. Each POP may supply Internet connections to one or more users and servers. The connection between POPs, users, and servers may include any suitable transmission media, including, but not limited to, public telephone lines, T1 lines, T3 lines, dial-up, DSL (Digital Subscriber Line), cable, Ethernet or wireless connections. The computers may be connected over a network such as the Internet, an intranet, a wide area network (WAN), local area network (LAN), or any other type of network. The computers may also be directly connected to one another or any number of other user computers. The computer may be a client computer coupled to an Internet service provider over a SLIP (Serial Line Interface Protocol) or PPP (Point to Point Protocol) connection. The Internet service provider is, in turn, coupled to the Internet, the client computer thereby having the ability to send and receive information to other nodes on the Internet using a TCP/IP protocol (Transmission Control Protocol/Internet Protocol).

It is to be understood that the network configuration and interconnections shown in Fig. 1 and described herein, are provided for purposes of illustration only. One of ordinary skill in the art will readily appreciate that the present invention may be practiced on networks more or less complex than that shown, in accordance with the teachings contained herein.

Fig. 2 illustrates additional detail of the SPAM reporting system 40. The system includes a SPAM detector 42, network address detector 44, SPAM database 46, host identifier 48, and report generator 50. The SPAM detector 42 may be any device configured for distinguishing SPAM e-mail from legitimate e-mail. In one embodiment, the detector is an intelligent SPAM detection system using statistical analysis or an updateable neural analysis engine, as disclosed in U.S. Patent Application Serial Numbers 09/916,599 and 09/916,930, both filed July 26, 2001, which are incorporated herein by reference in their entirety. These devices use a statistical analyzer to gather statistics associated with text in the e-mail message and a neural network engine coupled to the statistical analyzer which is taught to recognize unwanted messages based on statistical indicators. The statistical indicators are analyzed utilizing the neural network engine for determining whether the electronic mail message is an unwanted message. It is to be understood that other types of SPAM detectors may be used without departing from the scope of the invention.

10072708.020502

E-mail messages that are identified as SPAM by SPAM detector 42 are sent to network address detector 44, which is used to identify the URL (Uniform Resource Locator), or other applicable network address, of a web site pertaining to the message. As previously discussed, much of the information that is included in the SPAM e-mail message is typically spoofed, and therefore cannot be used to identify the true source of the mail. However, it is common for SPAM e-mails to contain URL's of web sites relating to the e-mail message. These are typically valid web sites since they must allow the recipient of the e-mail to follow up on the spammer's offer. The web site may contain, for example, information on how to obtain products or sign up for services advertised in the spammer's original message. The URL may then be used to track the origin of the spammer's e-mail or a web site they are using to sell their product or service.

10037203 .020502

In addition to locating the URLs within the e-mail, the network address detector 44 is configured to examine the text surrounding the URL to determine the likelihood that the URL is an address of the spammer's web site. For example, text within a SPAM e-mail may include:

> "Visit our web site at http:// . . ."; or
>
> "Come and see Sexy Suzy at http:// . . .".

The network address detector 44 is coupled to SPAM database 46 which contains common words or phrases associated with an advertised web site. The database

46 preferably uses wildcards in validating the surroundings of a URL as the spammer's web site.

The SPAM database 46 also includes a list of known valid (or trusted) senders of e-mails to rule out network addresses that may be present in the legitimate e-mail messages. For example, in the case where a SPAM e-mail was forwarded through an innocent party. The database may be pre-populated, but is preferably updateable by a system administrator to ensure that the SPAM reporting system 40 does not become a nuisance to innocent third parties.

10072708 . 020502

Once a network address is identified, host identifier 48 is used to locate the web server hosting the spammer's web pages. Many Internet service providers require their subscribers to sign contracts that forbid SPAM. It is therefore appropriate to report the SPAM to any service provider whose users originate SPAM. WHOIS, NSlookup, Finger, Telnet, Ping, Traceroute, or any other address tracing tool may be used to identify the ISP and report the problem. NSlookup allows for recovery of the IP address from a domain name. Traceroute demonstrates the route that a packet takes from an arbitrary Internet site to another arbitrary site.

If the URL contains a raw IP address, a reverse DNS (Domain Name Server) lookup may be used to identify the domain name of the web site. Once the domain name is found, a WHOIS lookup may be used to identify the individuals who are involved in maintaining the spammer's Internet domain. The WHOIS report contains various administrative contacts for the owner of the domain, such as shown below:

WHOIS Information for someorg.com

Registrar: NETWORK SOLUTIONS, INC.

Organization: Some Organization, Inc.
Address: 123 Some Lane, Somewhere

Admin contact: Hostmaster
E-mail: j.spammer@lotsaSpamISP.com
Phone: 123-4567
Fax: 876-5432

Tech contact: Hostmaster
E-mail: j.spammer@lotsaSpamISP.com
Phone: 123-4567
Fax: 876-5432

Nameservers: dns1.someorg.com
dns2.someorg.com

It also specifies the organization that the domain is registered with, and where this individual or organization's mail is hosted. This information is used by the report generator 50 to generate an e-mail message to the responsible organization incorporating details of the suspected SPAM, as further described below.

The WHOIS report may also contain additional contact information for parent organizations. For example, if a small ISP is hosted by a larger backbone provider this information may be included in the report. The system administrator may have the option of notifying the organization only, or also notifying the parent organization. As used herein, the term 'hosting authority' refers to any organization responsible, either directly or indirectly, for hosting the spammer's web site, domain, or e-mail account.

10072302.020502

The report generator 50 uses the hostmaster or postmaster e-mail address provided by the host identifier 48 to generate a report 52 which is sent by e-mail to the hosting authority. The report 52 may include, for example, content of the suspected SPAM e-mail, date and time the e-mail arrived on recipient's server, IP address and name reported during the SMTP connection, and the full WHOIS report used to track down the responsible authority. The IP address and name reported during SMTP connection may be spoofed, but this may be useful in tracking down an open SPAM relay that the spammer is using. The report 52 may also include disclaimer information and user definable text. The e-mail message used to transmit the report 52 to the relevant authority may also be signed to verify the source. It is to be understood that the report may contain less information than noted above or additional information without departing from the scope of the invention.

In order to prevent the SPAM reporting system 40 from becoming a nuisance to the authorities, the system 40 may include a device which restricts the frequency and number of reports sent to any given authority. For example, the information on spammers may be collected and reported only once a month.

The system 40 may also be configured to include one or more central Managed Service Providers (MSPs) which are responsible for collecting information from a number of organizations. Each MSP is responsible for reporting spammers to authorities once enough evidence has been collected from one or more organizations for a particular SPAM threat. The device reduces the chance of multiple organizations sending individual reports, and thus further reduces the possibility of the SPAM reporting system 40 becoming a nuisance itself.

The computer on which the SPAM reporting system is installed may be a stand-alone desktop computer, laptop computer, server, mainframe, or a mobile or handheld computing device (e.g., personal digital assistant (PDA) or mobile phone), for example. Fig. 3 shows a system block diagram of computer system 60 that may be used as the user computer, server, or other computer system to execute software of an embodiment of the invention. As shown in Fig. 3, the computer system 60 includes memory 62 which can be utilized to store and retrieve software programs incorporating computer code that implements aspects of the invention, data for use with the invention, and the like. Exemplary computer readable storage media include CD-ROM, floppy disk, tape, flash memory, system memory, and hard drive. Additionally, a data signal embodied in a carrier wave (e.g., in a network including the Internet) may be the computer readable storage medium. Computer system 60 further includes subsystems such as a central processor 64, fixed storage 66 (e.g., hard drive), removable storage 68 (e.g., CD-ROM drive), and one or more network interfaces 70. Other computer systems suitable for use with the invention may include additional or fewer subsystems. For example, computer system 60 may include more than one processor 64 (i.e., a multi-processor system) or a cache memory. The computer system 60 may also include a display, keyboard, and mouse (not shown) for use as a desktop or laptop computer.

10072808.020502

The system bus architecture of computer system 60 is represented by arrows 72 in Fig. 3. However, these arrows are only illustrative of one possible interconnection scheme serving to link the subsystems. For example, a local bus may be utilized to connect the central processor 64 to the system memory 62. Computer system 60 shown in Fig. 3 is only one example of a computer system suitable for use with the invention. Other computer architectures having different configurations of subsystems may also be utilized. Communication between computers within the network is made possible with the use of communication protocols, which govern how computers exchange information over a network, as is well known by those skilled in the art.

Fig. 4 is a flowchart illustrating a process of the present invention for identifying the source of a SPAM e-mail message and generating a report to notify the relevant authority. At step 80 an e-mail is received. The SPAM detector 42 determines whether the e-mail is legitimate or unsolicited (step 82). If the e-mail is found to be unsolicited the message is examined to find the network address of the web site relating to the message (step 84). Once the network address is identified, the host identifier 48 is used to locate the web server hosting the web site (step 86). Report 52 is then generated containing the information found by the host identifier 48 and additional information such as content of message, IP address, and information used to trace the message to the host (step 88). If an MSP is assigned to collect and send generated reports, the report is sent to the MSP, which in turn forwards the report to the appropriate authority (steps 90, 92, and 94). If an MSP is not assigned, but the system is configured to hold the reports for a period of time (e.g., collect all reports until the end of the month), the report is temporarily saved (steps 96 and 98). If the system is not configured to collect and hold reports, the report is sent immediately to the relevant authority (step 100).

Although the present invention has been described in accordance with the embodiments shown, one of ordinary skill in the art will readily recognize that there could be variations made to the embodiments without departing from the scope of the present invention. Accordingly, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

10072708.020502